DOCUMENT RESUME

ED 278 679 TM 870 096

AUTHOR Noble, Christopher S.; And Others
TITLE Omit Rates on Criterion-Referenced Tests for Different Ethnic Groups: Implications for Large Scale Assessment.
PUB DATE Apr 86
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Basic Skills; Blacks; *Criterion Referenced Tests; *Ethnic Groups; Goodness of Fit; High Schools; Hispanic Americans; *Item Analysis; Mathematics Tests; Minimum Competency Testing; Reading Tests; Regression (Statistics); Response Style (Tests); State Programs; *Test Bias; Testing Programs; Test Items; Test Length; *Test Theory; Whites
IDENTIFIERS *Omitted Responses; Texas Assessment of Basic Skills; *Texas Educational Assessment of Minimum Skills

ABSTRACT

The relationship between item omission and item position on criterion-referenced tests in the Texas state assessment program is examined. Item statistics from the Texas Educational Assessment of Minimum Skills (TEAMS) and Texas Assessment of Basic Skills (TABS) mathematics and reading tests from 1983 through 1985 are examined for three ethnic groups and two grade levels. Sample sizes ranged from approximately 15,000 for the TEAMS field test to about 225,000 for the regular TABS testing program. Ethnic representation was approximately proportionate to that of the state. On these untimed, basic skills assessment instruments, the rate at which students omitted any item appears to increase in a regular manner with item position. The rate of increase of tiem omissions, as well as the omit rate itself, were both found to differ significantly for students from different ethnic groups. Simple regression procedures were used to fit prediction equations to these data in order to model the rate of omission as a function of item analysis. The results of this study provide information for framing test administration procedures in such a way that students from any particular ethnic group are not unfairly penalized. Categorized data are illustrated in tables and figures following the study. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278679

Omit Rates on Criterion-Referenced Tests for Different Ethnic Groups: Implications for Large Scale Assessment

Division of Student Assessment
Texas Education Agency
Austin, Texas

Christopher S. Noble
Gail D. Nicholson
Edgar N. Morgan

April 11, 1986

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. S. Noble

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

★ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 870 096

Abstract

The relationship between item omission and item position on criterion-referenced tests in the Texas state assessment program is examined. Item statistics from the Texas Educational Assessment of Mimimum Skills (TEAMS) and Texas Assessment of Basic Skills (TABS) mathematics and reading tests from 1983 through 1985 are examined for three ethnic groups and two grade levels. Ethnic representation was approximately proportionate to that of the state. On these untimed, basic skills assessment instruments, the rate at which students omitted any item appears to increase in a regular manner with item position. The rate of increase of item omissions as well as the omit rate itself, were both found to differ significantly for students from different ethnic groups. The results of this study provide information for framing test administration procedures, in such a way that students from any particular ethnic group are not unfairly penalized.

Omit Rates on Criterion-Referenced Tests for Different Ethnic Groups: Implications for Large Scale Assessment

Summary

The relationship between item omission and item position on criterion-referenced tests in the Texas state assessment program is examined. Item statistics from the Texas Educational Assessment of Minimum Skills (TEAMS) and Texas Assessment of Basic Skills (TABS) mathematics and reading tests from 1983 through 1985 are examined for three ethnic groups and two grade levels. Sample sizes ranged from approximately 15,000 for the TEAMS field test to about 225,000 for the regular TABS testing program. Ethnic representation was approximately proportionate to that of the State of Texas.

On these untimed, basic skills assessment instruments, the rate at which students omitted any item appears to increase in a regular manner with item position. The rate of increase of item omissions as well as the omit rate itself, were both found to differ significantly for students from different ethnic groups.

Simple regression procedures were used to fit prediction equations to these data in order to model the rate of omission as a function of item position. Prediction models with different rates of increase of the student omit rate, as well as different omit rates for any particular item on the test, were obtained for students from different ethnic groups for both subject areas in each of the four test administrations held before October, 1985. Good fits were obtained with simple polynomial models for all four of these test administrations.

The results of this study provide information for framing test administration procedures, in such a way that students from any particular ethnic group are not unfairly penalized.

Purpose of Study

As a result of mandates from the Texas State Legislature, the Texas Education Agency has been conducting a statewide minimum skills assessment program since 1980. Criterion-referenced tests in mathematics and language arts have been administered at five different grade levels. Beginning in 1987, all Texas students will be required to pass the exit level tests in order to receive a high school diploma. Thus, in order to ensure that each student has a fair opportunity to receive a high school diploma, all feasible efforts must be made to eliminate any ethnic bias from the testing procedure (see AERA Guidelines, 1985).

This study illustrates how the tendency to omit items on untimed tests appears to follow a regular pattern with item position. This tendency increases with test length while the specific pattern varies with ethnicity. As a result, students from some ethnic groups may be

unfairly penalized, particularly in regard to their performance on items towards the end of such tests, by improper administration of unusually long tests.

A review of the literature has not produced any previous studies that examine the relationship between omit rate and test length by ethnic group. The tendency to omit items has, however, been described previously as a deviant response characteristic (Hattie, 1983).

The purpose of this study is to provide information for the framing of test administration procedures in such a way that effective steps can be taken to eliminate this source of potential ethnic bias.

Methodology and Data Sources

The item level data from criterion-referenced mathematics and reading tests given to Texas high school students throughout the state from 1983 to 1985 were analyzed for students from three different ethnic groups. Specifically, black, Hispanic, and white student results were examined. These tests were administered as a part of the TABS (1982-1985) and the TEAMS (1985) statewide testing programs. Data were examined for more than 200,000 students who took the TABS tests at Grade 9, as well as for 14,500 Grade 11 students who took the 1985 TEAMS field test, and for 200,000 Grade 11 students who took the 1985 TEAMS regular administration. The relevant characteristics of the populations are described in Table 1. These data were used to investigate the relationship between the percent of students omitting each item on the test and the location of that item in sequence on the test. Data on student performance, measured as the percent of students mastering each item for these administrations, were also examined. Multiple forms of the test using different sequences of the items were administered. Items of varying difficulty were randomly arranged within each test and form in the 1983 and 1985 TABS regular administrations and in the 1985 TEAMS field test. A single form was used in the 1984 TABS and 1985 TEAMS regular administrations. There was no increase in item difficulty with position in the tests. Also all test items were judged to be free from ethnic bias when the tests were constructed. Item bias statistics may be obtained from the Texas Education Agency. Finally, data were also examined from one statewide testing session after administration procedures were changed to limit situations in which students might omit items. This administration also included increased motivation since the Grade 11 TEAMS mathematics and language arts tests given October 1985 were tied to high school graduation.

Analysis and Results

In each case (i.e., for each test administration, for each ethnic group, for each subject area) the percent of students omitting each item on the test was plotted against the location of that item in sequence on the test. Simple linear regression procedures utilizing non-linear

transformations were used to fit prediction equations to these data in order to model the student omit rate as a function of item position.

Good fits were obtained with simple polynomial (cubic) functions (Figure 1 and Table 2). Specifically, more than 90% of the variance was explained in many cases. For those cases where less variance (65% to 88%) was explained by the fitted model, individual items with unusually high difficulty levels appeared to be causing this distortion. As such, student omit rates for these items should be considered outliers for the purposes of this study. For each test administration and subject area, three different cubic relations modeling this trend were established for the three different ethnic groups being considered. The parameters in these three trend models differed significantly from each other and from zero at a significance level of less than 0.0001. Relevant statistics are summarized in Table 2.

In each case, the shape of the above mentioned relationship, as reflected in these plots and models, was such that there appeared to be:

(i) an approximately constant omit rate for each ethnic group for the first 20 items on the tests,

(ii) a significant increase in this omit rate between the 20th and 30th items on the tests,

(iii) an approximately constant (though, in actuality, slightly increasing) rate of increase in the student omit rate with item position, after about the 30th item, in sequence, on the tests, and

(iv) significantly different student omit rates as well as different rates of increase of the student omit rate for different ethnic groups, within the same administration.

Accordingly, in each case, the approximate item number at which this significant increase in the student omit rate occurred was estimated. Also estimated for each case, was the average rate of increase of this student omit rate for those items located in sequence, after the item at which the above mentioned significant increase occurred. The effects of these different omit rates, in terms of the increased number of students from each ethnic group omitting the final item on each test, were also documented. Both predicted and actual values were recorded and compared. These values are summarized in Table 3.

The relationship between item position and student performance, measured as the percent of students correctly answering each item on the test (or p-value) was plotted for all cases. Examples are shown in Figures 2 and 3. Analysis of this relationship revealed no significant trend. Finally, patterns of omit rates were examined for the results of the statewide TEAMS test conducted in October 1985.

For this test, administration procedures had been specifically modified to reduce the significance of the omit phenomenon. As shown in Figure 4, there was no relationship between item position and the omit rate for this administration.

Major Findings

The results of the analyses indicate that, for four statewide test administrations from February 1983 to May 1985, the student omit rate begins to increase significantly, in a regular manner, between the 20th and 30th items in sequence on each test. This holds for students from three ethnic groups in both subject areas considered. However, for the statewide test administration held in October 1985, this phenomenon was not present. At this last testing, different administration procedures had been implemented to ensure that all students, both in their perception and in actuality, were given sufficient time to respond to all questions on the test. Although these were untimed tests, it is thought that some students were not given sufficient time on the earlier administrations. As a result, some students did not even guess answers for many items. This is evidenced by the trends of the omitting patterns discussed above, as well as by the slightly higher omit rate for some items at the beginning of each test. Students were not penalized for incorrect answers on these tests; rather, students had a 1 in 4 chance of correctly guessing the answer to an item. Student test taking strategies usually involve completing the answers to easier items regardless of their position in the test, before answering the more difficult items, some of which may be early in the test. Clearly, a large number of students were either not encouraged or not given sufficient time to complete these items.

For the four earlier administrations, the most important factor affecting the student omit rate appears to be ethnicity, if we exclude individual items with very low p-values as outliers. In general, the trend of the student omit rate with item position was to increase in proportion to the third power of the item position, across all of these four earlier administrations, for students from all three ethnic groups and for two subject areas. However, the cubic equations modeling this trend for students from the various ethnic groups taking each test incorporated significantly different modifying parameters. This reflects the fact that the omit rate as well as the rate of change with item position of this omit rate differ significantly for different ethnic groups. Since this omit rate increases in proportion to the third power of the item position, the number of students omitting each item rises very quickly to a significant size, with increasing item number. Moreover, since these student omit rates increase at significantly different rates for different ethnic groups, test length becomes crucial when measuring overall test performance. For example, the fraction of students omitting the final item in each area on a 44 item test ranged from a low of 1.4% to a high of 4% for different ethnic groups. Specifically, 1180 black students (4.0%), 1810 Hispanic students (3.2%), and 1720 white students (1.4%) omitted this item on the 1983 TABS administration. Had the test been longer, the discrepancies

between these percentages would likely have been larger for every succeeding item. See Table 3 and Figure 1.

By contrast, for the lone testing with the revised administration procedures, the numbers and percentages of students omitting the 72nd and final item on the mathematics test were 38 (0.16%), 50 (0.11%), and 93 (0.08%) for black, Hispanic, and white students, respectively. On all items in this test, omitting patterns correlated only with item difficulty, as measured by p-value, for both subject areas. See Table 3 and Figures 3 and 4.

Conclusion and Recommendations

It is clear that students with higher omit rates could be unfairly penalized on untimed tests, and that this penalty increases on longer tests with more difficult items. Since students from different ethnic groups tend to omit items at different rates, it is also clear that students from the groups with higher rates would be penalized unfairly as a group. Accordingly, procedures to eliminate this ethnic bias from future administrations of such tests need to be considered and implemented. These preventive measures were taken in Texas for the statewide administration of a high school minimum skills test for over 200,000 students. The results of this single administration indicated that the regular pattern and increasing proportion of omits, with item position, were not present. The only significant correlation with omit rates was for item difficulty (as measured by p-value) as would be expected in an untimed criterion-referenced test. See Figures 3 and 4.

The lack of pattern for this last administration indicates that the phenomenon of differential student omit rates on untimed tests does not reflect differential knowledge, but rather bias introduced by the testing procedure.

The observation of slightly higher omit rates for all ethnic groups at the beginning of each of the earlier statewide administrations also rules out different test taking strategies as a factor in this pattern. In fact, this indicates that all students used similar strategies and that a number of students from all ethnic groups were either not motivated or not permitted time to return to uncompleted items.

Implications with regard to the causes and results of the previously observed effects need to be investigated further. Hattie (1983) mentions that "distortions could be considerable if omits are present" and that "there are a variety of possible reasons" related to fatigue. Madsen and Murray (1984) point out that "high anxiety students were troubled by factors such as time constraints and exam length." Results of our study further indicate that certain populations may be even more severely penalized whenever these factors are overlooked.

References

AERA Guidelines for Eliminating Race and Sex Bias in Educational Research and Evaluation, Educational Researcher, June/July 1985, p.16-17.

Hattie, John. (1983) "The Tendency to Omit Items: Another Deviant Response Characteristic." Educational and Psychological Measurement, 43, p.1041-1045.

Madsen, Harold S. and Murray, Norma. (1984) "Retrospective Evaluation of Testing in ESL Content and Skills Courses." Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages, Houston, TX., March 6-11, 1984.

TABLE 1
Sample Characteristics

Year	Program	Grade	Subject	Number of Students Tested			
				Black	Hispanic	White	Total[1]
1983	TABS[2]	9	Mathematics	29581	56519	122657	212797
			Reading	29416	56599	121841	211895
1984	TABS	9	Mathematics	29911	58955	129723	222810
			Reading	29517	58669	128519	220907
1985	TABS	9	Mathematics	30791	60558	133351	229391
			Reading	30421	59955	132115	227140
1985	TEAMS[3] Field Test	11	Mathematics and Reading[4]	2150	3450	8650	14500
1985	TEAMS	11	Mathematics	23917	44691	114754	187360
			Language Arts	23741	44463	114370	164875

1 Includes a very small number of Asian and American Indian students.
2 Texas Assessment of Basic Skills
3 Texas Assessment of Minimum Skills
4 Estimated values (± 100)

TABLE 2
Goodness of Fit Statistics for Cubic Prediction Equation
(Omit Rate v $[\text{Item Number}]^3$)

Year	Program	Grade	Subject	R^2 Value (% of variance explained by model) Black	Hispanic	White
1983	TABS	9	Mathematics	92	92	91
			Reading	86	84	80
1984	TABS	9	Mathematics	71	70	79
			Reading	82	84	81
1985	TABS	9	Mathematics	92	95	93
			Reading	95	94	91
1985	TEAMS Field Test	11	Mathematics	84	88	77
			Reading	87	92	65
1985	TEAMS	11	Mathematics	N.A. (no reasonable fits)		
			Language Arts	N.A. (no reasonable fits)		

TABLE 3
Summary of Significant Student Omit Rates

Year	Program	Subject	Test Length (# of items)	Ethnic Group	Omit Rate at Final Item (%)[2]		Item # at Significant Increase in Omit Rate (Fatigue Point)	Average Rate of Change of Omit Rate after Fatigue Point (% omits/item)
					Actual	Predicted		
1983	TABS[1]	Math	44	Black	4.0	5.1	22	0.19
				Hispanic	3.2	3.8	22	0.15
				White	1.4	1.7	22	0.07
		Reading	44	Black	3.1	2.5	26	0.10
				Hispanic	2.4	1.9	26	0.09
				White	0.8	0.6	26	0.02
1984	TABS	Math	44	Black	2.6	2.7	29	0.12
				Hispanic	1.7	1.9	29	0.08
				White	0.8	0.9	29	0.04
		Reading	44	Black	1.5	1.7	23	0.12
				Hispanic	1.0	1.2	24	0.03
				White	0.4	0.4	23	0.01
1985	TABS[1]	Math	50	Black	4.1	3.7	21	0.10
				Hispanic	4.1	3.8	21	0.11
				White	1.6	1.5	22	0.04
		Reading	54	Black	3.3	3.1	24	0.08
				Hispanic	3.2	2.9	24	0.08
				White	1.0	0.9	24	0.02
1985	TEAMS[1] Field Test	Math	45	Black	17.2	21.3	20	0.73
				Hispanic	11.9	12.3	20	0.45
				White	3.3	3.8	20	0.14
		Reading	45	Black	12.0	13.4	20	0.46
				Hispanic	7.4	7.7	20	0.28
				White	1.8	1.9	20	0.08
1985	TEAMS	Math	72	Black	0.2	n.a.	(no reasonable fits)	
				Hispanic	0.1	n.a.	(no reasonable fits)	
				White	0.1	n.a.	(no reasonable fits)	
		Language Arts	72	Black	0.2	n.a.	(no reasonable fits)	
				Hispanic	0.2	n.a.	(no reasonable fits)	
				White	0.1	n.a.	(no reasonable fits)	

[1] Actual omit rates given are the averages over several forms of the test.
[2] % of students omitting this item.

EXAMPLE OF OMIT RATE TRENDS AND P-VALUE WITH ITEM POSITION (TABS FEBRUARY 1983)

Figure 1

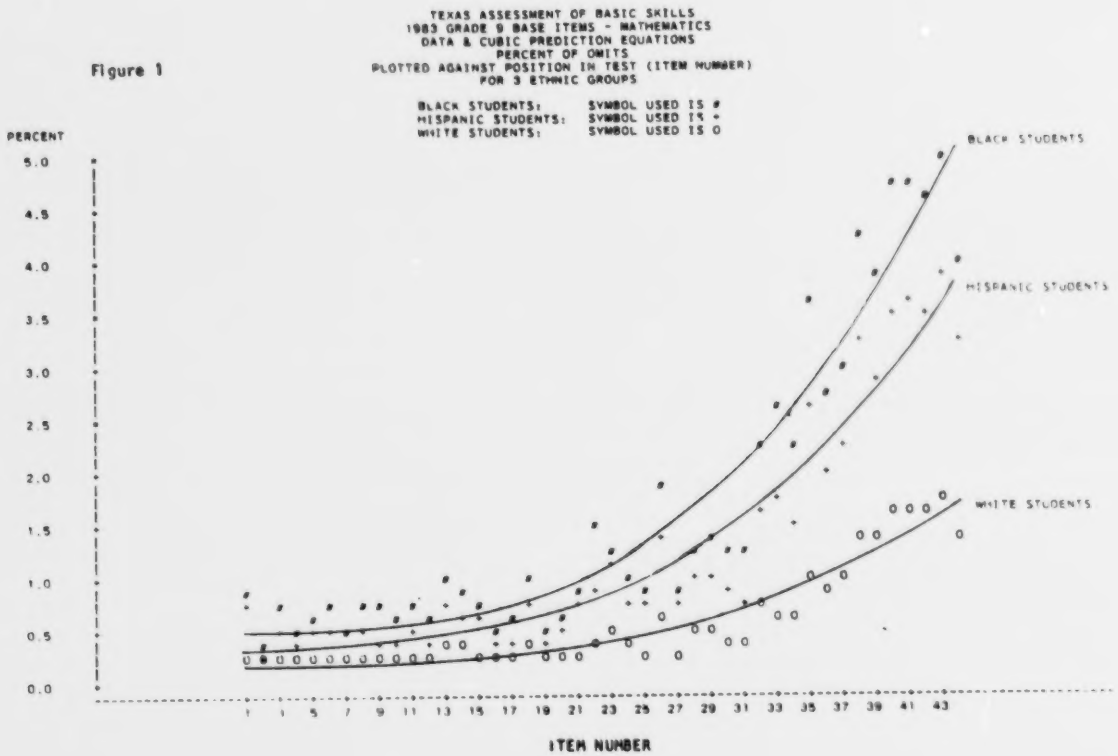
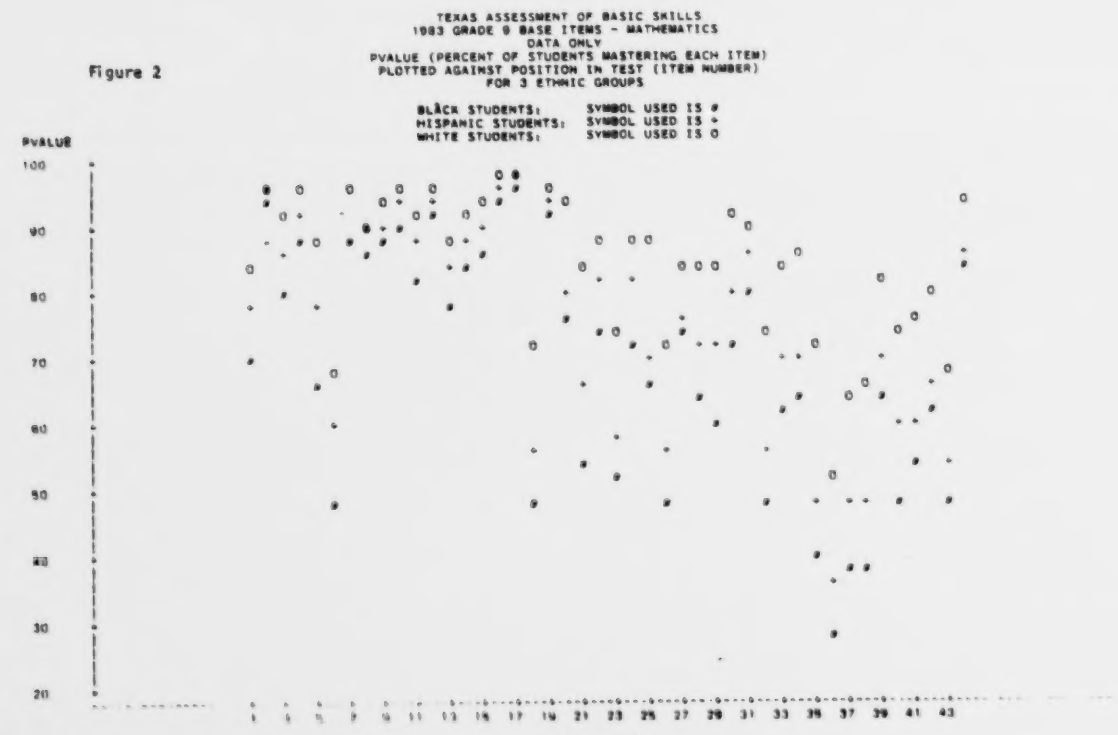


Figure 2


TEXAS ASSESSMENT OF BASIC SKILLS
1983 GRADE 9 BASE ITEMS - MATHEMATICS
DATA ONLY
PVALUE (PERCENT OF STUDENTS MASTERING EACH ITEM)
PLOTTED AGAINST POSITION IN TEST (ITEM NUMBER)
FOR 3 ETHNIC GROUPS
BLACK STUDENTS: SYMBOL USED IS #
HISPANIC STUDENTS: SYMBOL USED IS +
WHITE STUDENTS: SYMBOL USED IS 0
PVALUE
100
90
80
70
60
50
40
30
20
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43
ITEM NUMBER

EXAMPLE OF TREND FREE OMIT RATES AND P-VALUES
(TEAMS OCTOBER 1985)

Figure 3

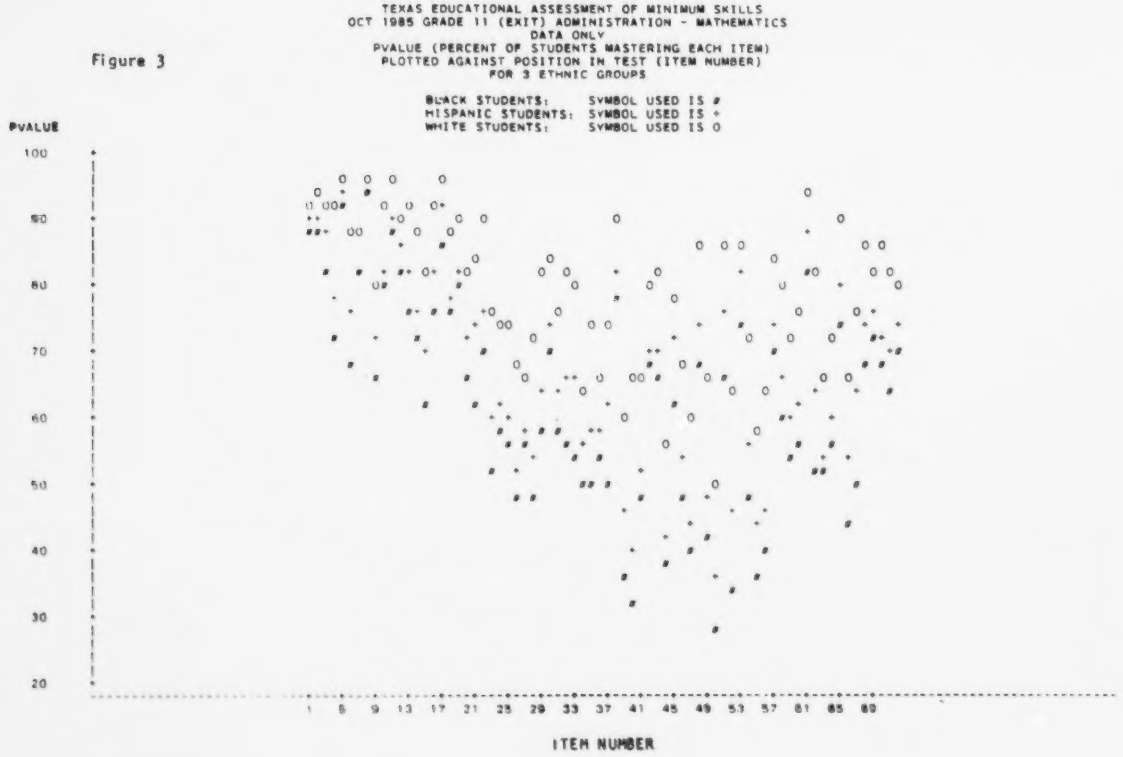
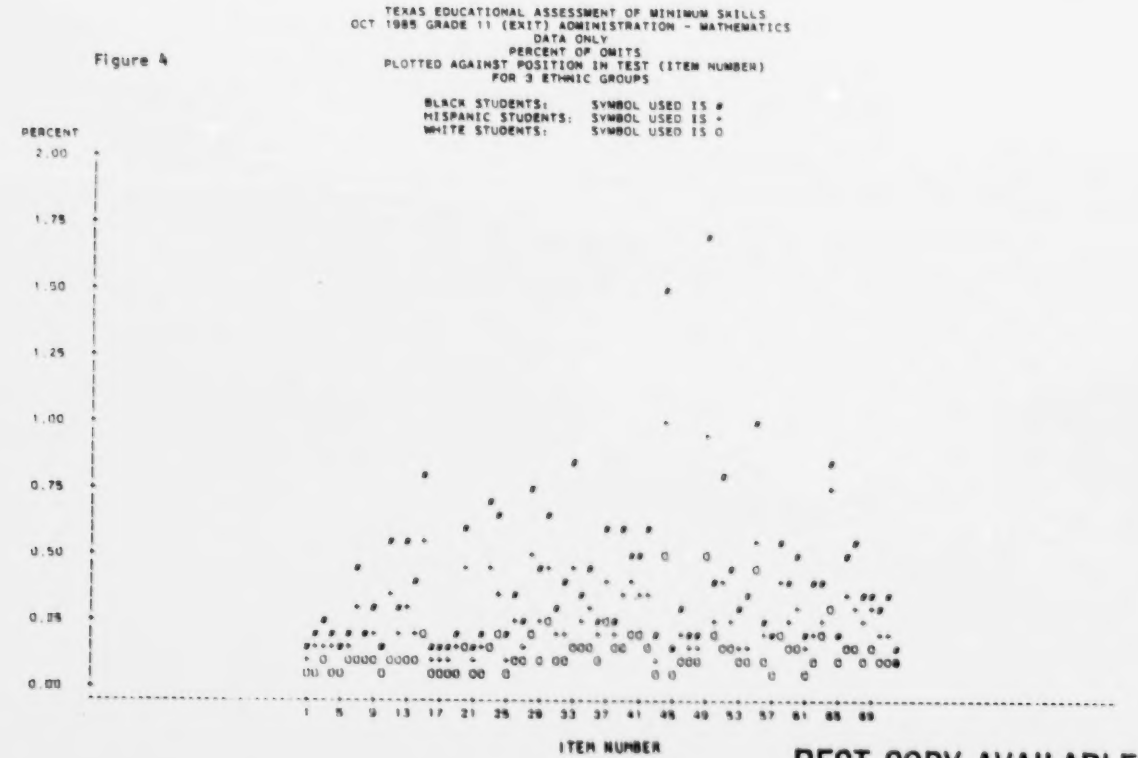


Figure 4



BEST COPY AVAILABLE